Advanced Data Acquisition -

Task 1

Western Governor’s University

Performance Assessment

Matthew Morgan

Student ID: 010471280

5/30/2023

A1:

For this project I used the WGU medical data set which is already loaded into the Postgres server in the virtual lab. The other data set I used is the [2013-2014 NHANES data provided by the CDC](https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey) from Kaggle.com. All the data in this data set is included in the submission.

A2:

To load the dashboard I included all necessary files in a zip file that was part of the submission. In the real world this zip file could be provided through e-mail or in a shared drive that the receiving party has access to such as OneDrive, Google Drive, or any other file sharing platform. To install the dashboard follow the steps below:

1. Download D211MorganFiles.zip to your computer.
2. Right click on the zip file, select “Extract All” and extract the files to the following file path ‘C:\Users\Public\Public Downloads’
3. On the desktop, open pgAdmin by double clicking on the shortcut.
4. In the left hand pane, click on the arrow next to servers to expand it.
5. Now right-click on medical\_data and select Query Tool near the bottom of the menu that pops up.
6. Once the query tool opens you will want to select the open file button. This looks like a manilla folder that is right above the new query tool pane on the right hand side.
7. Navigate to ‘C:\Users\Public\Downloads’ and select the sql query.txt file that was included in the original zip file.
8. If it doesn’t show up then you will have to choose “All Files” in the drop down above the select button in the bottom right hand corner.
9. Once the sql query.txt file has been loaded, hit the run button above the query tool pane. This looks like a play button.
10. Once the query is complete, you will get a confirmation message, close pgAdmin, choose ‘Leave’ if prompted.
11. Navigate back to the ‘C:\Users\Public\Public Downloads’ and find the file titled ‘D211MorganWorkbook’.
12. Double-click on it.
13. When prompted, choose “Skip This Update.”
14. Once loaded you will be prompted to enter a username and password. Use the following:
    1. Username: postgres
    2. Password: Passw0rd!
    3. Notice the 0 is the number zero.
15. The dashboard has now been loaded and you can navigate it or manipulate it however you want.

A3:

After successfully loading the dashboard, select the ‘Story 1’ tab on the bottom if the file doesn’t default to that view. To enter presentation mode you can hit F7 on your keyboard, or select the bottom on the toolbar that looks like a projector screen. Once in presentation mode you will have a fullscreen view of the dashboard that I have created for this project.

The story in this project is a singular dashboard created based on comparable data between the CDC and WGU datasets. I used similar visualizations as part of my project from D210 but tried to simplify them for this project.

To navigate this dashboard there are many options. Each individual visualization can be used to filter the rest of the dashboard, this allows you to filter for gender by clicking on the correct slice in the pie chart, filter for a specific age bin from each source in the bar chart, select a specific source in the check box filter, filter for patients from WGU/CDC data sets who have been diagnosed with Asthma/Diabetes/Overweight, as well as those patients who have been diagnosed in specific age bins by clicking on the appropriate cell in the tables provided. This allows us to look at overall trends of patients from both sources, gender, age, and diagnosis.

A4.

A .txt file of all the SQL code has been uploaded as part of the submission.

B:

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=794e1333-333d-40ee-adaa-b01201438fb9>

C1:

Throughout this program, the primary goal of the WGU dataset is to try and reduce patient re-admissions. In an effort to help accomplish that goal, I have used Tableau to try and compare the WGU vs the CDC patient demographics. When comparing the data sets through D210 and D211 it’s obvious that the WGU patients are more likely to be older, overweight, diabetic, and asthmatic. I tried to create visualizations that showed the disparity in which this occurs.

The reason I feel this is important is that the patient population is effected by adverse health conditions at a higher rate than CDC collected data. This is having an adverse effect on patient re-admissions. If the patient population can be treated properly to help those numbers regress to the mean, we’ll probably have better patient outcomes.

C2:

Tableau was used as that was the program available to use in the virtual lab, as well as being a very common analysis and visualization tool used in many businesses worldwide. The features of programs like Tableau allow for easy accessibility from other stakeholders within organizations to allow for further introspection and analysis.

C3:

The following steps were used to prepare the data for analysis in Tableau.

1. Explore already loaded tables within pgAdmin.
2. I noticed that foreign keys weren’t assigned so I wrote code to do so.
3. Use CAST to change datatypes in the WGU dataset. (FROM Yes/No TO True/False Boolean strings)
4. Adding a source column to the data set with a WGU label so it can be differentiated within Tableau.
5. Assign “Non-Binary” designation to “Prefer not to answer” responses in gender in accordance to the data dictionary.
6. Set max age of 80 to WGU data to remain consistent with CDC data that records any person over 80 as 80.
7. Create tables for CDC questionnaire and demographics csv files.
8. Import CSV files into created tables.
9. Convert data in needed columns in accordance with WGU data. (1 = True, 2 = False)
10. Add a source column to the CDC data set with a CDC label to it can be differentiated within Tableau.
11. Remove all minors (Age less than 18) from both data sets as WGU doesn’t have any minor data.
12. Create temporary tables for WGU (wgunew) and CDC data (nhanesnew) with only relevant columns to be used in Tableau.
13. Use UNION ALL to combine all columns from the temporary tables to create one final\_data table that will be used in Tableau.
14. Use DROP TABLE IF EXISTS on both temporary tables so prevent any errors from occurring if tables don’t exist.

By using the UNION command, more advanced analysis can be done within Tableau since all data is in one table. This eliminates the need to create a Join in Tableau where there is no column to Join on.

C4:

Below are directions on creating each worksheet, dashboard, and story in Tableau.

**Connect Tableau to Server for Data**

1. Open Tableau
2. Under Connect > To A Server > More…
3. Select PostgreSQL and enter the following information
   1. Server: localhost
   2. Port: 5432
   3. Database: medical\_data
   4. Authentication: Username and Password
   5. Username: postgres
   6. Password: Passw0rd!
4. Click Sign In
5. Drag “final\_data” table to the pane where it says “Drag tables here”

Now click on Sheet 1 on the bottom

**Age Bin Bar Graph**

1. Open the drop down on Age in the Data Pane
2. Select Create > Bins
3. In size of bins, type 5
4. Click OK
5. Drag the newly created Age(Bin) to columns
6. Drag source to columns
7. Drag original Age to rows
8. It will default to sum, but what we want is count. In the drop down select Measure > Count
9. Drag source to color in Marks pane.
10. In the visualization do the following:
    1. Right click on Sheet 2 and select Edit Title, replace <Sheet Name> with the text ‘Age Comparison’ and center it.
    2. Right click on Age (bin) / Source and select Hide Field Labels for Columns
11. Rename sheet to ‘Age Comparison’

**Asthma % Table**

1. Click on “New Worksheet”
2. Drag source to columns
3. Drag Asthma to rows
4. Drag Asthma to Text box in Marks pane
5. Select Measure > Count within the Asthma marks drop down
6. Select Quick Table Calculation > Percent of Total within the Asthma marks drop down
7. Select Compute Using > Table (Down) within the Asthma marks drop down
8. On the visualization do the following:
   1. Right click on sheet 1 and select hide title
   2. Right click on False and select hide
   3. Right click on Asthma and select Hide Field Labels for Rows
   4. Right click on True and deselect Show Header
   5. Right click on source and select Hide Field Labels for Columns
9. Rename sheet to ‘Asthma Population’

**Asthma Count Table**

1. Click on “New Worksheet”
2. Drag Age (bin) to columns
3. Drag source to rows
4. Duplicate this sheet twice
5. Continue on sheet 3
6. Drag Asthma to Text box in Marks pane
7. In the same pane, select Measure > Count in the Asthma drop down.
8. On the visualization do the following:
   1. Right click Sheet 3 and select hide title
   2. Right click Age (bin) and select Hide Field Labels for Columns
   3. Left click on 15 and select exclude
9. Rename Sheet to ‘Asthma Panel’

**Diabetes % Table**

1. Click on “New Worksheet”
2. Drag source to columns
3. Drag Diabetes to rows
4. Drag Diabetes to Text box in Marks pane
5. Select Measures > Count within the Diabetes marks drop down
6. Select Quick Table Calculation > Percent of Total within the Diabetes marks drop down
7. Select Compute Using > Table (Down) within the Diabetes marks drop down
8. On the visualization do the following:
   1. Right click on sheet 6 and select hide title
   2. Right click on False and select hide
   3. Right click on Diabetes and select Hide Field Labels for Rows
   4. Right click on True and deselect Show Header
   5. Right click on source and select Hide Field Labels for Columns
9. Rename sheet to ‘Diabetes Population’

**Diabetes Count Table**

1. Select Sheet 3(3)
2. Drag Diabetes to Text box in Marks pane
3. In the same pane, select Measure > Count in the Diabetes drop down.
4. On the visualization do the following:
   1. Right click Sheet 3 and select hide title
   2. Right click Age (bin) and select Hide Field Labels for Columns
   3. Left click on 15 and select exclude
5. Rename Sheet to ‘Diabetes Panel’

**Overweight % Table**

1. Click on “New Worksheet”
2. Drag source to columns
3. Drag overweight to rows
4. Drag overweight to Text box in Marks pane
5. Select Measures > Count within the overweight marks drop down
6. Select Quick Table Calculation > Percent of Total within the overweight marks drop down
7. Select Compute Using > Table (Down) within the overweight marks drop down
8. On the visualization do the following:
   1. Right click on sheet 7 and select hide title
   2. Right click on False and select hide
   3. Right click on overweight and select Hide Field Labels for Rows
   4. Right click on True and deselect Show Header
   5. Right click on source and select Hide Field Labels for Columns
9. Rename sheet to ‘Overweight Population’

**Overweight Count Table**

1. Select Sheet 3(2)
2. Drag Overweight to Text box in Marks pane
3. In the same pane, select Measure > Count in the Overweight drop down.
4. On the visualization do the following:
   1. Right click Sheet 3 (2) and select hide title
   2. Right click Age (bin) and select Hide Field Labels for Columns
   3. Left click on 15 and select exclude
5. Rename Sheet to ‘Overweight Panel’

**Gender Pie Chart**

1. Click on “New Worksheet”
2. Drag Gender to color in the Marks pane
3. Drag Gender to angle in the Marks pane
4. Under the Gender Angle Marks drop down select Measure > Count
5. In the main drop down in the marks pane select Pie
6. Drag Source into the Filters pane
7. In the filter pop up select All and click OK
8. In the Source Filter drop down select Show Filter
9. In the Source Filter drop down on the right hand side select Customize > Show Apply Button
10. In the visualization right-click on sheet 8 and select Edit Title…
11. Delete the current text and type in ‘Distribution of Gender’
12. Rename the sheet to Gender Pie Chart

**Dashboard**

1. Click on “New Dashboard”
2. In the bottom left, below the objects pane, select floating
3. Drag Age Comparison towards the middle of the top of the sheet
4. Select the Legend pane that pops up and click on X to remove it
5. Re-size age comparison bar chart
6. Drag Gender Pie Chart to the left of the Age Comparison Bar Graph
7. Drag the Pie Chart Legend to the bottom of the pie chart
8. Drag the source filter to the top right of the dashboard
9. Click on the Age Comparison bar graph, select “Use as Filter”
10. Click on the Pie Chart and select “Use as Filter”
11. Drag the Asthma Population and Asthma Panel to be next to each other below the Pie Chart and Bar Chart
12. Do the same for the Diabetes worksheets and place below the Asthma visualizations
13. Do the same for the Overweight worksheets and place below the Diabetes visualizations
14. Right click on select hide title on each of the Asthma, Diabetes, and Overweight visualizations
15. Select “Use as Filter” on each of the Asthma, Diabetes, and Overweight visualizations
16. In the objects Pane drag a text box to the top of each set of Asthma, Diabetes, and Overweight Visualizations. Title each box appropriately. Make the text centered, size 24, and Tableau Bold font. Label each section “x Population Statistics” where x is Asthma, Diabetes, or Overweight.
17. Make sure to make the text box takes up the width of the dashboard so it’s centered appropriately.
18. Finally, on the source filter drop down hover over Apply to Worksheet, and select Selected Worksheets…, then check the box next to worksheets to select all the worksheets in the workbook. Click OK.

**Story**

1. Click on “New Story”
2. Drag Dashboard 1 into the “Drag a sheet here” pane
3. Navigate the dashboard as needed

C5:

There are a few takeaways from the data visualization that I built. Those being that the WGU patient population is largely older and less healthy than that of the real US population based on CDC data. WGU patients are up to three times as likely to have Asthma, Diabetes, or be overweight.

In looking at the big picture, this disproportionate occurrence of these health issues has to have some effect on our patient re-admission rate. Why this is the case, or what the WGU hospital system can do for these patients needs to be researched and looked into in-depth.

C6:

There are a number of limitations in this data analysis. These are listed below:

1. The CDC data is purported to be a snapshot of the entire United States population. Therefore, it’s representative of people who are generally healthy and not in the hospital much, if at all. Whereas the WGU data is a snapshot of patients who are currently in the hospital and whether they got re-admitted or not in the future. This could be a big reason why the WGU patient population is older and more sickly.
2. This also means that the comparisons made in my visualizations aren’t a great basis to make huge decisions on. It can help us to notice trends and provide a basis to begin case studies, it’s not a solid foundation to make large scale changes. Again, this is due to the fact that the CDC data is based on a random sampling of the United State population whereas the WGU data is based on patients who are currently in the hospital.
3. Gender comparisons can’t be made appropriately between these two data sets. The CDC data set wasn’t encoded to non-binary genders whereas the WGU dataset provided a non-binary option for patient identification.